A neural-based vocoder implementation for evaluating cochlear implant coding strategies

El Boghdady, Nawal ; Kegel, Andrea ; Lai, Waikong ; Dillier, Norbert

Abstract: Most simulations of cochlear implant (CI) coding strategies rely on standard vocoders that are based on purely signal processing techniques. However, these models neither account for various biophysical phenomena, such as neural stochasticity and refractoriness, nor for effects of electrical stimulation, such as spectral smearing as a function of stimulus intensity. In this paper, a neural model that accounts for stochastic firing, parasitic spread of excitation across neuron populations, and neuronal refractoriness, was developed and augmented as a preprocessing stage for a standard 22-channel noise-band vocoder. This model was used to subjectively and objectively assess consonant discrimination in commercial and experimental coding strategies. Stimuli consisting of consonant-vowel (CV) and vowel-consonant-vowel (VCV) tokens were processed by either the Advanced Combination Encoder (ACE) or the Excitability Controlled Coding (ECC) strategies, and later resynthesized to audio using the aforementioned vocoder model. Baseline performance was measured using unprocessed versions of the speech tokens. Behavioural responses were collected from seven normal hearing (NH) volunteers, while EEG data were recorded from five NH participants. Psychophysical results indicate that while there may be a difference in consonant perception between the two tested coding strategies, mismatch negativity (MMN) waveforms do not show any marked trends in CV or VCV contrast discrimination.

DOI: https://doi.org/10.1016/j.heares.2016.01.005

A Neural-Based Vocoder Implementation for Evaluating Cochlear Implant Coding Strategies[☆]

Nawal El Boghdady[a,*], Andrea Kegel[b], Wai Kong Lai[b], Norbert Dillier[b]

[a] *Institute for Neuroinformatics (INI), Universität Zürich (UZH)/ ETH Zürich (ETHZ), Zürich, Switzerland*
[b] *Laboratory of Experimental Audiology, ENT Department, Universitätsspital Zürich (USZ), Zürich, Switzerland*

Most simulations of cochlear implant (CI) coding strategies rely on standard vocoders that are based on purely signal processing techniques. However, these models neither account for various biophysical phenomena, such as neural stochasticity and refractoriness, nor for effects of electrical stimulation, such as spectral smearing as a function of stimulus intensity. In this paper, a neural model that accounts for stochastic firing, parasitic spread of excitation across neuron populations, and neuronal refractoriness, was developed and augmented as a preprocessing stage for a standard 22-channel noise-band vocoder. This model was used to subjectively and objectively assess consonant discrimination in commercial and experimental coding strategies.

Stimuli consisting of consonant-vowel (CV) and vowel-consonant-vowel (VCV) tokens were processed by either the Advanced Combination Encoder (ACE) or the Excitability Controlled Coding (ECC) strategies, and later resynthesized to audio using the aforementioned vocoder model. Baseline performance was measured using unprocessed versions of the speech tokens.

Behavioural responses were collected from seven normal hearing (NH) volunteers, while EEG data were recorded from five NH participants. Psychophysical results indicate that while there may be a difference in consonant perception between the two tested coding strategies, mismatch negativity (MMN) waveforms do not show any marked trends in CV or VCV contrast discrimination.

Keywords: Neural vocoder, ACE, ECC, objective measures, EEG, MMN.

1. Introduction

In a typical Nucleus device, processing an incoming sound signal usually involves a Fast Fourier Transform (FFT) filterbank stage that decomposes the acoustic signal into 22 spectral channels, corresponding to the total number of implanted electrodes. The standard coding strategy used for these devices is the Advanced Combination Encoder (ACE), which selects only the n channels with the highest energy content from the available 22. These n electrodes are stimulated with biphasic pulses at the predefined channel stimulation rate. In this coding strategy, the intensity of the acoustic signal is encoded in the amplitude of the stimulating biphasic pulses (Zeng et al., 2008). However, the ACE strategy accounts for neither the refractory period of the auditory nerves nor the electric field interaction between neighbouring electrodes.

Encoding loudness information in the amplitude of the stimulating pulses leads to louder signals that have poorer spectral resolution due to increasing electric field interactions between electrodes. One way

[☆] This work was part of the first author's master thesis in partial fulfilment of the program requirements at the Institute for Neuroinformatics (INI), Universität Zürich (UZH)/ ETH Zürich (ETHZ), Switzerland.

[☆☆] Present address of corresponding author: Department of Otorhinolaryngology, Head and Neck Surgery, University Medical Center Groningen, Groningen, The Netherlands
Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Groningen, The Netherlands

[*] Corresponding author
Email address: n.el.boghdady@umcg.nl (Nawal El Boghdady)

to address this problem is to instead encode the incoming acoustic signal's intensity levels in the channel stimulation rate while keeping the pulse amplitude constant (Lai and Dillier, 2012). In this scenario, the pulse amplitude is kept at the threshold level of audibility (or slightly above threshold), while the stimulation rate varies as a function of the intensity of the original sound signal. Such an approach is implemented in the Excitability Controlled Coding (ECC) algorithm, which is a custom-designed coding strategy that takes into account the parasitic electrode interactions and neuronal refractoriness in addition to encoding the signal intensity in the stimulation rate (Babacan et al., 2010).

The perceptual differences between ACE and ECC have not been fully explored yet. For instance, compared to ACE, ECC is more likely to spread the resultant activity across the electrode array, similar to the PACE coding strategy (Nogueira et al., 2005). Any resultant reduction in perceived loudness due to the reduced stimulus density could be compensated for by increasing the stimulation level, in the same way this is accounted for with the PACE strategy (Büchner et al., 2008). In terms of speech discrimination, possible benefits of neurophysiologically based techniques, such as ECC, may be manifest in improved identification of Consonant-Vowel (CV) transitions. Psychoacoustic or electrophysiological discrimination of voice onset time and formant transitions may be studied using synthesized speech tokens (Klatt, 1980) with variations of specific phonological features in discrete steps (Raz and Noffsinger, 1985; Iverson, 2001; Stephens and Holt, 2011; Digeser et al., 2009; Henkin et al., 2008; Hant and Alwan, 2000).

To evaluate different coding strategies, especially during the development stages, vocoder simulations of the processed sound signal can be presented to normal hearing (NH) listeners (Shannon et al., 1995; Fu and Shannon, 1999; Lai et al., 2003; Strydom and Hanekom, 2011; Chen and Loizou, 2011; Massida et al., 2011; Chen, 2012). These signals are hypothesized to simulate speech cues transmitted through the implant. Such listening tests with NH subjects are often performed to allow optimization of algorithm parameters in addition to the identification of potential problems that might occur during the processing stages.

One problem with existing vocoder implementations (e.g. Strydom and Hanekom, 2011; Chen and Loizou, 2011; Chen, 2012; Massida et al., 2011) is that they are based on purely signal processing concepts, and hence do not take into account important biophysical phenomena involved with stimulus perception in cochlear implant (CI) users. For this reason, a vocoder implementation based on a neural model is advantageous, since it helps more closely simulate the cues perceived by CI subjects than a typical vocoder implementation.

Psychoacoustic experiments may be time-consuming and attention-dependent, thus it is beneficial to develop evaluation procedures based on objective measures, such as Event-Related Potentials (ERP)s that can be obtained from raw Electroencephalography (EEG) recordings (Lonka et al., 2013; Kraus et al., 1993). ERPs have been successfully recorded for both CI and NH subjects (Kraus et al., 1993) and may be useful for revealing discrimination abilities for various speech (Kraus et al., 1993; Kühnis et al., 2013) and music features (Sandmann et al., 2010; Lonka et al., 2013; Zhang et al., 2013; Agrawal et al., 2013).

Thus, the aim of this study is to develop a platform for objectively testing the output of various coding strategies in terms of consonant discrimination in different vowel contexts (CV or Vowel-Consonant-Vowel (VCV)). A neural vocoder model based on biophysical parameters was implemented to simulate the outputs of both ACE and ECC. Two experiments were then carried out: in Experiment I, these simulations of ACE and ECC were used to psychophysically test speech perception with NH volunteers. In Experiment II, EEG data was collected from a subset of those NH subjects using the Optimum-I oddball paradigm (Näätänen et al., 2004). Mismatch Negativity (MMN) waveforms were then calculated from the raw EEG data for different CV and VCV contrasts. Possible correlations between the psychophysical and EEG data were investigated.

2. Materials and Methods

2.1. Vocoder Processing

2.1.1. Standard Vocoder

All stimuli were processed via a program developed by (Omran et al., 2010; Laneau et al., 2006) based on modules from the Nucleus Matlab Toolbox (NMT) provided by Cochlear. Stimuli were loaded and processed once by ACE and once by ECC using a standard test map with the following parameters: channel

stimulation rate was set to 900 Hz with 10 maxima, 22 electrodes (channels), and threshold and comfort levels at 0 and 100 Current Level (CL) units, respectively, for all electrodes. The CL scale is a clinical current scale used to describe the intensity of a pulse, and ranges between the threshold and comfortably loud levels (T- and C-levels, respectively) that vary between electrodes and across patients. The CL scale is related to μA according to the following equation:

$$I_{stim} = 17.5\mu A * 100^{CL/255} \tag{1}$$

Processing WAV files with the designated coding strategy yields pulse sequences for each channel that can be used to stimulate a CI patient's electrode array. This sequence is usually represented in the form of a channel-time matrix, in which the pulse magnitude on each of the 22 channels is shown versus time (Figure 1) (Lai and Dillier, 2013).

These pulse sequences were resynthesized back into an audio signal using a standard noise-band vocoder. The envelope of each channel was used to modulate a noise signal with the same frequency band, and then these outputs were summed across all channels to yield the reconstructed audio signal.

Such a vocoder implementation is unsuitable for resynthesizing audio from pulses generated by ECC because it assumes that loudness information is encoded in the pulse amplitudes but not in the pulse rates, leading to an almost inaudible output. Furthermore, merely amplifying this audio signal introduces distortions, which renders the signal inappropriate for further testing. In addition, this vocoder does not take into account any biophysical phenomena, such as parasitic spread of excitation and neuronal refractoriness. To address these limitations, a neural model was implemented as a preprocessing stage to the standard vocoder.

2.1.2. Neural-based Vocoder Model

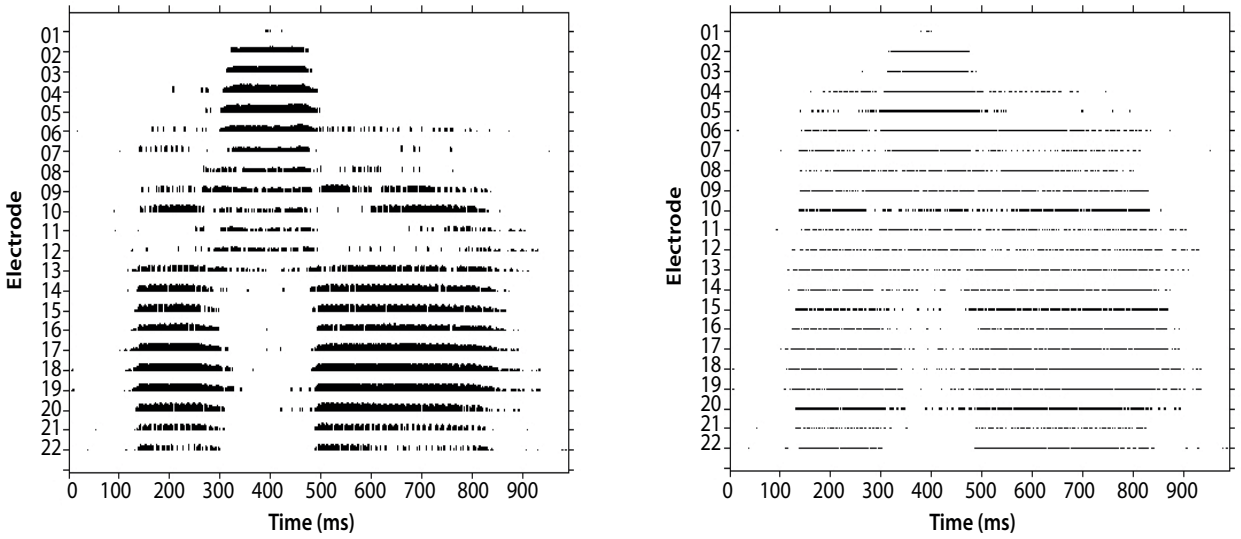
The neural model stage, which takes an arbitrary pulse sequence and processes it using a simple neural network, was based largely on the work in (Bruce et al., 1999b,c,a; McKay and McDermott, 1998; Cohen et al., 2003; Cohen, 2009a,b,c,d,e; Chen and Zhang, 2007; Florentine and Zwicker, 1979; McKay et al., 2003). In these studies, various neuronal models were proposed to account for loudness perception as a function of pulse rate and pulse amplitude. Various aspects of these models were integrated together and modified to have three processing stages, similar to those in Bruce et al. (1999b); Fredelake (2012); Fredelake and Hohmann (2012); Hamacher (2004), in order to obtain a block that takes a CI pulse sequence as input and produces a response which mimics that of a typical Auditory Nerve (AN) fibre when stimulated using the same pulse sequence.

In the first stage of the proposed model, Integrate and Fire (I&F) neuron populations are stimulated with the designated pulse sequence (e.g. Figure 1). Depending on the amplitude of the stimulating current, the influence of an electrode may spread to neighbouring neuronal populations as demonstrated in Figure 2 by the coloured curves (Cohen et al., 2003; Lai and Dillier, 2012). For example, low-amplitude stimuli, as indicated by the red pulses in Figure 2, produce smaller spatial spread compared to higher amplitude stimuli (the blue pulses). If the amplitude of the stimulus is very large, then its effect may spread to other populations that should normally be stimulated by neighbouring electrodes. The green pulses show that when the rate of stimulation is high, the neurons are stimulated strongly. The low amplitude of the individual green pulses helps limit the spread of excitation from affecting neighbouring populations. The output of this stage is the spiking activity (0 or 1) for each neuron versus time. Please note that adaptation is not included in the proposed model.

In the second stage, the weighted sum of the output spikes of each I&F neuron population is averaged by a single spatiotemporal integrator (Figure 2), to yield the average population activity. Altogether, there are 22 spatiotemporal integrator units corresponding to the number of channels.

In the third stage, the average population activity per channel versus time is then grouped into a matrix whose entries are passed through a loudness scaling function. The final output of this phase is a matrix whose entries are scaled between 0 and 100 on the CL scale. This matrix resembles a spectrogram (Figure 4), such that the "loudness" information is shown for each frequency band (channel) with time.

Each stage of the model is described in detail in the following sections.



(a) ACE (b) ECC

Figure 1: Examples of a typical pulse sequence for processing the token /a-Sa/. The height of each pulse represents the stimulus amplitude re. to the CL scale. The variable stimulation rate in ECC is used to encode loudness. These electrodograms were generated by a modified version of the software described in Lai and Dillier (2013).

Leaky Integrate and Fire Neurons

. The leaky I&F neuron model is described by Equation 2, which is the finite difference approximation of the equation provided in (Chen and Zhang, 2007). V_{mem} is the neuron's membrane voltage as a function of time, Δt represents 1/stimulation rate, V_{rest} is the neuron's resting membrane potential, R is the channel resistance, I_{stim} represents the positive phase of the biphasic stimulus current in Ampere as a function of time (Fredelake, 2012), and τ is the neuron's RC time constant, where C is the neuron's inherent membrane capacitance. The term V_{noise} is normally distributed between ± 2 mV, which is added to introduce stochastic firing behaviour.

In this I&F model, the neuron acts as an integrator which builds up its membrane voltage, V_{mem}, according to the input current's magnitude and timing, in addition to the neuron's R and C values. If V_{mem} exceeds the neuron's inherent threshold voltage (V_{thr}), the neuron fires an action potential, or spike, resets V_{mem} back to the resting potential value, V_{rest}, and the neuron enters a refractory state. During this absolute refractory period, the neuron cannot respond to any novel stimulus (Chen and Zhang, 2007).

$$V_{mem}(t+\Delta t) = V_{mem}(t) + \frac{\Delta t}{\tau}\Big[-(V_{mem}(t) - V_{rest}) + RI_{stim}\Big] + V_{noise} \tag{2}$$

Overall, 1000 I&F neuronal instances were created to simulate the AN fibres. Increasing the number of AN fibres beyond 1000 is not expected to improve the quality of the resynthesized speech (Holmberg et al., 2007).

Pulse sequences that were used to stimulate the I&F neuron populations were first converted from the CL scale to μA using Equation 1. Negm and Bruce (2014) provide a model for a single node of Ranvier in a mammalian AN fibre, with values of R and C given to be 1953.49 MΩ, and 0.0714 pF, respectively. Each of the 1000 neurons was assigned an R value normally distributed between 1900 MΩ and 2000 MΩ, and a C value normally distributed between 0.07 pF and 0.5 pF. Additionally, each neuron was assigned a random V_{rest} value normally distributed between -80 mV and -55 mV, and a random V_{thr} value normally distributed between -50 mV and -36 mV, according to electrophysiological measurements made in Waters

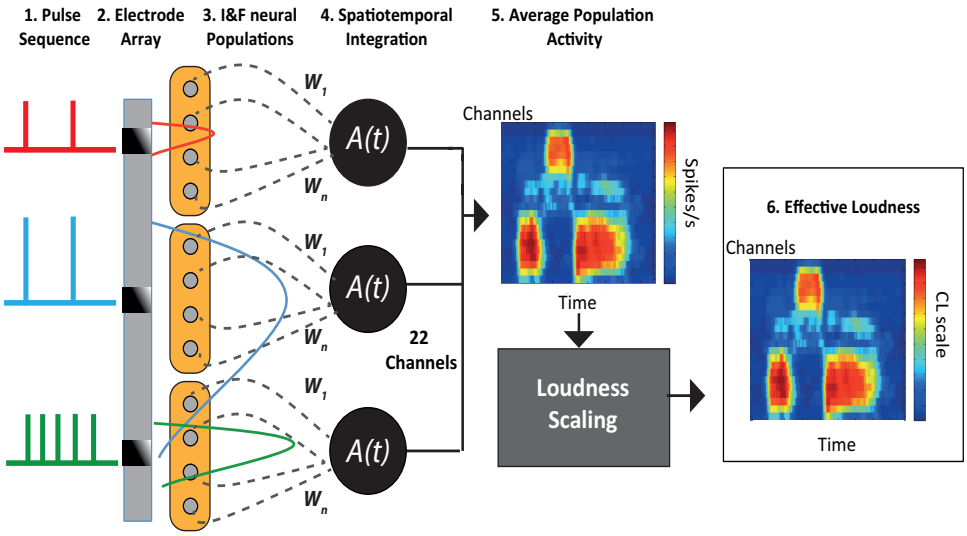


Figure 2: Neural model stages. The pulse sequence arrives as input to the electrode array. Each electrode is assigned an I&F neural population consisting of roughly 48 neurons. The coloured curves represent the current spread between neighbouring neural populations depending on the stimulus current intensity. The output spikes of each neural population are spatiotemporally averaged to obtain the population activity per channel versus time. The integration function $A(t)$ is defined by Equation 4. The population activity is then scaled between the T- and C-levels.

and Helmchen (2006). Note that in this paper, the authors recorded from pyramidal neurons in the cortex of anesthesized rats. The exact parameter values for the I&F neurons are given in Table 1.

Each electrode was assigned a population of 47-48 neurons as represented by the light grey circles in Figure 2. It is assumed that each electrode is placed such that maximal stimulation lies right in the centre of its corresponding neural population. This is unrealistic, however it significantly simplifies computation.

The general trend of the spread profile was elegantly modelled in Cohen et al. (2003) by a set of exponential curves, which demonstrate that as the stimulus current amplitude increases, both the width and the peak of the spatial spread profile also increase. The I&F neuron model incorporates Cohen et al. (2003)'s spatial spread profile, whereby the spatial spread is a Gaussian distribution with a mean μ and standard deviation σ. The peak, which is defined by the mean of the spatial spread distribution, is always placed at the centre of the neural population . The spatial spread constant, which is defined by the standard deviation σ, is set as a function of the stimulus current such that larger current pulses result in a wider spread. This is illustrated by Equation 3. The conversion factor of $10^{-3}(distance/\mu A)$ is essential for scaling down the value of the current to be suitable for use as a standard deviation. Otherwise, σ would be too large to produce an appropriate Gaussian curve, with its peak at a normalized value of *1*. In this model, the unit of distance is arbitrary, and represents the width of the spread curve from the centre of the neural population. The whole curve is then scaled according to the input current amplitude to account for the dependence of the peak on the stimulus level (Hartmann et al., 1984).

$$\sigma(I_{\mu A}) = I_{\mu A} * 10^{-3}_{\frac{distance}{\mu A}} \tag{3}$$

Each neuron is then stimulated with the current spread functions and if the neuron's membrane voltage, V_{mem}, exceeds its inherent threshold, V_{thr}, the neuron fires an action potential (spike), resets its membrane potential back to the resting potential, V_{rest}, and enters a refractory state. Each neuron was assigned a random absolute refractory period from a normal distribution between 1 μs and 300 μs. However, these values are less than the mean absolute refractory period (about 330 μs) measured from AN fibres in cats

Table 1: Parameter values for the I&F neuronal model.

Parameter	Value(s)	Reference
R	Normally distributed between 1900 MΩ and 2000 MΩ	(Negm and Bruce, 2014)
C	Normally distributed between 0.07 pF and 0.5 pF	(Negm and Bruce, 2014)
Absolute Refractory Period	Normally distributed between 1μs and 300 μs	mean of about 330μs in Miller et al. (2001)
Total number of I&F neurons used	1000	(Holmberg et al., 2007)
Number of neurons per population	$\sim$ 48	-
Number of populations	22	-
V_{thr}	Normally distributed between -50 mV and -36 mV	(Negm and Bruce, 2014; Waters and Helmchen, 2006)
V_{rest}	Normally distributed between -80 mV and -55 mV	(Negm and Bruce, 2014; Waters and Helmchen, 2006)

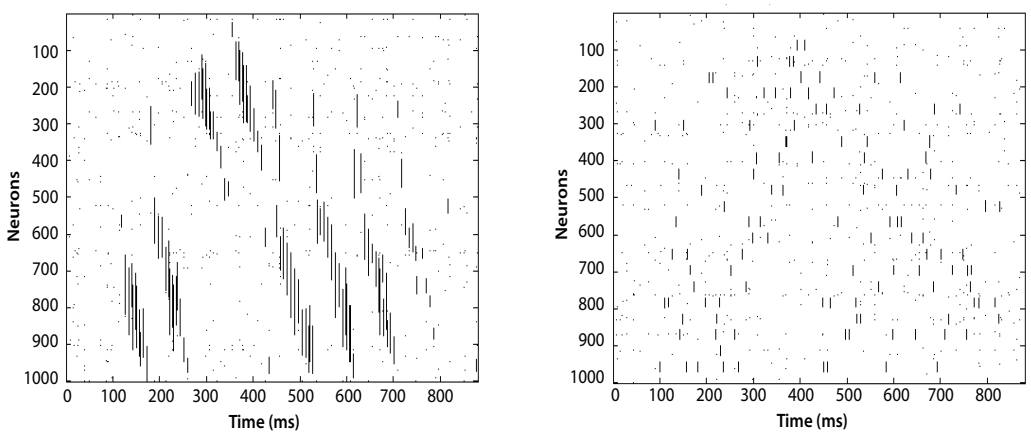
(Miller et al., 2001). Additionally, relative refractory periods were not included in this model. The output spike patterns of every neuron are then recorded in a raster matrix (Figure 3).

It should be noted that Adamson et al. (2002) provide some evidence for differences in firing behaviour between apical and basal AN fibres. Apical neurons exhibit slow adaptation properties and a prolonged latency compared with more basal neurons. Moreover, Fu (2005) found that loudness balance functions between apical and basal electrodes varied with stimulation rate. These results suggest that low frequency stimulation may be processed differently across different electrodes, which also varies from patient to patient. For simplicity, differences between the firing properties or responses of basal and apical neurons were not considered in this model.

Spatiotemporal Integration

. The spikes output by the I&F neuron layer are then relayed to an integrator unit in the second stage of the model. The 47-48 I&F neurons within each population are connected to a single integrator as shown in Figure 2, with the strength of each connection scaled by its synaptic weight. For computational simplicity, it is assumed that the synaptic weights $W_i \rightarrow W_n$ are normally distributed, with a peak value of *1* assigned at the middle of the population. This means that the centre-most neuron is always assumed to contribute the most to the spatial sum, and hence is assigned a maximum possible weight of 1. The weights of all other neurons in the periphery decrease following a Gaussian curve. These weights are then multiplied by the sum of the spikes output from each I&F neuron over a time window Δt, which is set to 35 ms to improve computational speed. This value is close to the time window duration of 10 ms specified in (McKay and McDermott, 1998). This Δt is a different parameter from the one used in Equation 2. Inhibition effects are not considered here, even though inhibitory neurons play an important role in higher level cortical computations.

The output of the integrator is the average activity of its corresponding I&F neuron population, as computed by Equation 4. $A(t)$ represents the average population activity as a function of time, n is the total number of neurons in a given population, Δt is the averaging time window, and W_i is the weight of the i^{th} neuron in the population. This equation is modified from the one described in Gerstner and Kistler (2002) in order to include synaptic weighting when considering the contribution of each neuron in the population. There are 22 integrators in total, representing the 22 channels. More importantly, applying

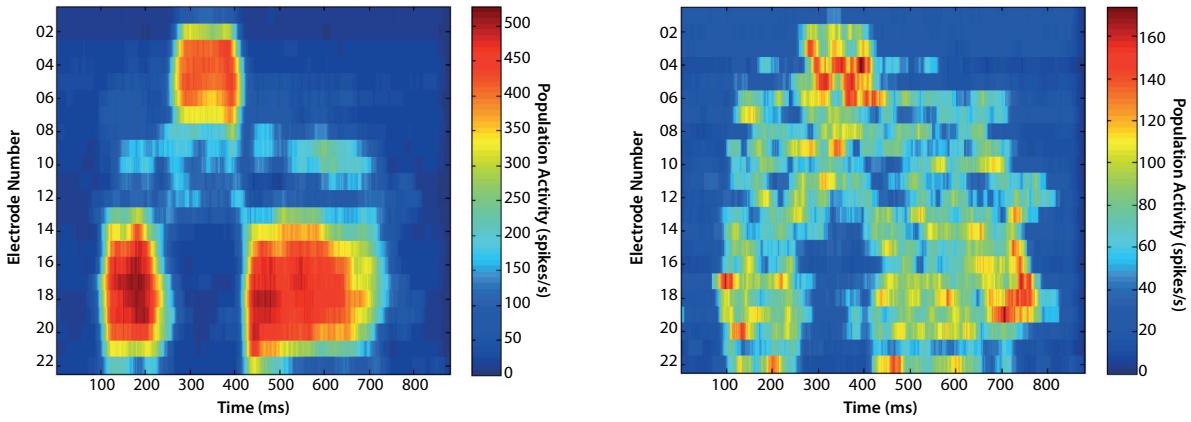


(a) ACE (b) ECC

Figure 3: Spike rasters for processing the speech token /a-Sa/. The intensity of the stimulus leads to a wider current spread across more neurons (represented on the vertical axis) in ACE compared to ECC. This is hypothesized to result in worse spectral resolution.

Equation 4 to the output of the I&F neurons converts the data from *spikes* to *spike rate*, which is a more meaningful measure that can be used in conjunction with psychophysical data to model loudness growth perception (McKay and McDermott, 1998; Yates et al., 1990). This output is shown in Figure 4.

$$A(t) = \frac{1}{n * \Delta t}\left[\sum_{i=1}^{n}\left(W_i \sum_{t}^{t+\Delta t} spikes_{i,t}\right)\right] \tag{4}$$



(a) ACE (b) ECC

Figure 4: Population activity matrix for processing the speech token /a-Sa/ with ACE (4(a)) and ECC (4(b)).

Loudness Estimation
. The output of the integrator layer has to be converted from *spike rate* to a suitable loudness scale before

resynthesis. The literature contains various models of the relationship between neuronal firing rate as a function of stimulus intensity in dB (Bruce et al., 1999a; Yates et al., 1990; Chatterjee and Zwislocki, 1998; Zwislocki, 1973). In these papers, a general sigmoidal function translating stimulus intensity to neuronal firing rate is defined, albeit with different parameter sets.

The sigmoid loudness growth function had still to be optimized for the model output, especially in setting the threshold. Following the example of Pasley et al. (2012), who used a linear model to reconstruct speech from spikes recorded from human cortex, a piece-wise linear approximation to the sigmoid was used, which allowed the slope and midpoint (and indirectly the threshold) of the function to be set independently of each other. The threshold of this function was empirically set to 1/4 the maximum firing rate among all neuronal populations. Any firing rate above this threshold was linearly transformed to a value between T- and C-level (0 and 100, respectively).

Greenwood Synthesis Filters

. Because cochleae vary in size and shape from one person to another, electrode locations are different across patients (Stakhovskaya et al., 2007). The stimulated tonotopic location can be approximated based on the length of the patient's cochlea and the electrode array insertion depth. Greenwood showed that a general tonotopic function can describe the frequency-to-place mapping in the cochleae of multiple species, including humans, if the size of the cochlea was scaled accordingly (Greenwood, 1990). Equation 5 shows this relationship, where for a human cochlea, $A = 165.4$ (to give a centre frequency F in Hz), $a = 0.06$ if the distance on the basilar membrane, x, measured from the apex is in millimeters, and $k = 1$ for humans (Greenwood, 1990).

The cochlear length and insertion depth parameters were set to the average dimensions of 33 mm and 22 mm, respectively, to approximate each electrode's location on the basilar membrane (Başkent and Shannon, 2004). The tonotopic frequencies corresponding to those locations were then used as the vocoder synthesis filters.

$$F_{Hz} = A(10^{ax} - k) \tag{5}$$

Note that this model assumes the stimulated frequencies along the Spiral Ganglion (SG) are identical to those along the organ of Corti. This is not the case, as was shown in Stakhovskaya et al. (2007), in which a function was derived to map the Greenwood frequencies along the organ of Corti to their corresponding tonotopic locations along the SG. Such a function would provide a more accurate model for the electrode-neural interface.

2.2. Stimuli

The use of CV or VCV speech tokens in consonant perception experiments, such as those utilized in (Raz and Noffsinger, 1985; Dorman et al., 1997; Stephens and Holt, 2011), arises from the observation that speech confusion in CI patients occurs mainly between stop consonants that differ in place of articulation. In this study, two sets of stimuli were used: a group of synthetic CV tokens, and another comprised of naturally-spoken VCV tokens.

There were four versions of each stimulus: an unprocessed version (referred to as *unprocessed*), a version that was processed by ACE and then resynthesized using the standard vocoder (ACE_{old}), another version also processed by ACE but resynthesized with the neural-based vocoder (ACE_{neural}), and a final version that was processed by ECC and resynthesized using the neural-based vocoder(ECC_{neural}).

2.2.1. CV tokens

For this stimulus set, 13 different stimuli were generated using a Klatt synthesizer (Klatt, 1980). Stimuli 1, 7, and 13 represent the *anchor* tokens /ba/, /da/, and /ga/ respectively, while the others provide the transitions between those tokens: from anchor /ba/ to anchor /da/ and from anchor /da/ to anchor /ga/. The term *anchor* refers to the tokens that were most distinguished as /ba/, /da/, and /ga/. The *transition* tokens on the other hand were artificial manipulations done to the anchor stimuli to achieve a full stimulus continuum. These transition tokens were created by varying the formant frequencies in discrete steps for the

consonant part of the tokens, as was done in Raz and Noffsinger (1985). For a single token, the total duration was set to 300 ms: The consonant transients lasted for 40 ms, while the steady-state vowel component lasted for 260 ms, as highlighted in Figure 5. Because the differences between these anchor tokens lie in only one or two formant transitions, CI processing with limited spectrotemporal resolution would lead to a high level of confusion among these tokens.

These stimuli were chosen because their consonants belong to the same phonetic group, the oral stop consonants, and thus are usually confused by CI subjects (Raz and Noffsinger, 1985). This makes them useful for investigating the degree of consonant confusion across different CI coding strategies.

Each token was processed using the four conditions (unprocessed, ACE_{old}, ACE_{neural}, and ECC_{neural}), which yielded 52 different stimuli (13 tokens * 4 processing conditions).

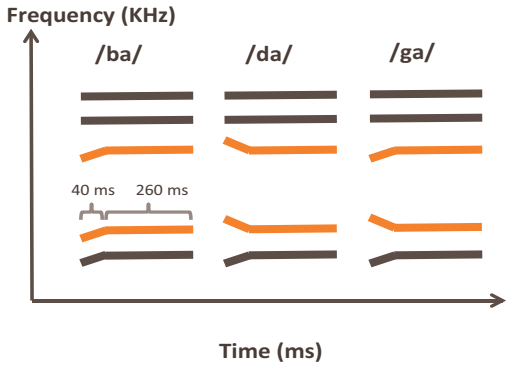


Figure 5: Schematized spectrograms of the unprocessed /ba/-/da/-/ga/ anchor tokens. Differences between those tokens lie in only the second and third formants. Figure inspired by stimuli representation provided in Hant and Alwan 2000.

2.2.2. VCV tokens

An additional VCV test battery was included. The speech tokens in this set were all prerecorded natural stimuli spoken by Jochen Sotscheck (Sotscheck, 1982). This stimulus group was chosen to compare subjects' performance on tests with synthetic versus natural stimuli.

There were 12 tokens in total. All tokens started and ended in the vowel **/a/**, with a consonant embedded in the middle from the group *b, d, g, f, k, l, m, n, p, r, s, t.* This gave rise to the following tokens: **/a-Ba/**, **/a-Da/**, **/a-Ga/**, **/a-Fa/**, **/a-Ka/**, **/a-La/**, **/a-Ma/**, **/a-Na/**, **/a-Pa/**, **/a-Ra/**, **/a-Sa/**, and **/a-Ta/**. The advantage of using this test set is that it not only allows examining subjects' performance on natural speech tokens, but also investigating discrimination difficulties among members of the sonorant and non-sonorant consonant groups.

2.3. Subjects

Seven NH participants (2 males) aged between 25 and 52 years (mean age 32.71 years; standard deviation 9.95 years) volunteered to take part in the psychophysical experiment. These volunteers had pure tone hearing thresholds on both ears of less than 20 dB HL from 250 Hz - 4 kHz at the time of testing. Only four of the seven participants were able to come back for the EEG experiment, thus one additional volunteer (male) was recruited for that experiment, yielding a total of five subjects. The participants were either fluent in German or had some working knowledge of that language. All subjects gave their informed consent prior to experimentation.

3. Experiment I: Psychophysics

3.1. Procedure

This experiment was conducted in a sound-proof anechoic chamber. Stimuli were presented from an insert-earphone placed in the subject's right ear, while the left ear was not stimulated. This was done to be

able to compare results obtained from the NH subjects tested in this experiment with those from unilaterally implanted CI subjects. The experimental protocol adopted here follows that defined in Raz and Noffsinger (1985); Stephens and Holt (2011); Iverson (2001).

3.1.1. CV Tokens

For this set, a three-alternative forced-choice task (3AFC) was used, in which the subjects listened to a random token from the 13 different files defining the /ba/-/da/-/ga/ continuum, and were then required to identify whether the token they heard was /ba/, /da/, or /ga/. Each token from the 13 was repeated 4 times for a total of 52 presentations in each of the four experimental conditions. In a given condition, all tokens were pseudorandomly shuffled and then presented to the participants.

All four experimental conditions (unprocessed, ACE_{old}, ACE_{neural}, and ECC_{neural}) were preceded with a training phase, in which the subjects were familiarised with the anchor tokens. This was followed by a self-test phase with feedback, again using only the anchor tokens, to familiarise subjects with the test procedure. Finally, the actual test was carried out, in which all 13 tokens were presented (anchor and transition), with 4 repetitions each and no feedback. The test was repeated, and the average results from the two tests were analysed. Each experimental session lasted for one hour, and subjects were asked to come back for a re-test session. In the re-test session, subjects also underwent training, self-test, and 2 actual test phases, and results from the 2 actual tests on the second session were averaged with those obtained from the first session.

3.1.2. VCV Tokens

A 12AFC task was used for this set, in which each token was presented 4 times in a pseudorandom order, for a total of 48 presentations per experimental condition. Each test was preceded with a training and a self-test phase, as with the CV set. All tests were conducted using the MACarena software (Lai and Dillier, 2002).

3.2. Results

3.2.1. CV Tokens

The percentage correct responses for each token were first averaged across the two test sessions, and then averaged across all participants for each experimental condition (Figure 6). The tokens presented are plotted on the x-axis, while the mean percentage responses for each token are shown on the y-axis.

The data show that the performance with processed tokens is poorer compared with the unprocessed condition, as indicated by the absence of well defined regions for each category in the processed conditions. For example, the ACE_{old} panel in Figure 6 shows a large confusion between the /da/ and /ga/ categories. From these data, the three anchor tokens (/ba/, /da/, and /ga/), in addition to the middle token on the /da/-/ga/ continuum, were chosen for the EEG experiment. The confusions for these four tokens are plotted in Figure 7 for better illustration. When CI processing is used, the cues for /ga/ become less salient since subjects were more likely to identify it as /da/ (the ACE_{old} panel in Figure 7). When the neural-based vocoder is introduced, those cues start emerging again compared to the ACE_{old} scenario, but at the expense of the cues for /ba/ (panels ACE_{neural} and ECC_{neural} in Figure 7).

3.2.2. VCV Tokens

Responses from each subject for this token group were first averaged across the two test sessions and then averaged across subjects. Subject responses for these tokens are represented as confusion matrices (Figure 8): the presented tokens are plotted on the x-axis, while the responses are shown on the y-axis. The third dimension corresponds to the mean percentage responses. For example, in the ACE_{old} panel in Figure 8, the token /a-Pa/ is correctly identified in 100% of the trials, while the token /a-Na/ is confused with /a-La/ almost 40% of the time. For this stimulus category, the tokens /a-La/, /a-Ba/, /a-Ma/, and /a-Na/ were chosen for the EEG experiment.

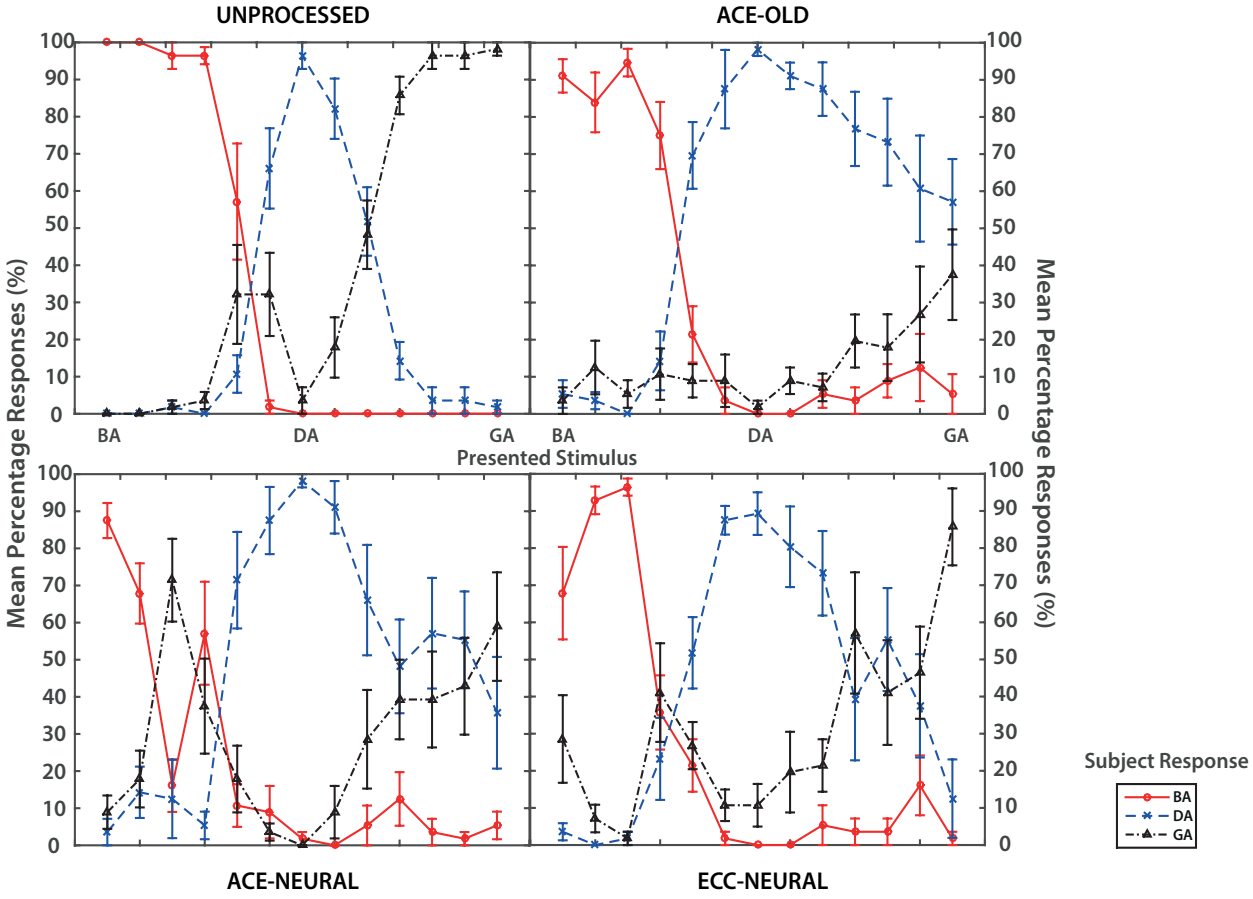


Figure 6: Mean percentage correct responses and standard errors for the CV set. Plot style inspired from (Raz and Noffsinger, 1985).

3.3. Discussion

3.3.1. CV Tokens

From Figure 5, one can see that the differences between /ba/ and the other two anchors lie in the second and third formants. However, the difference between the /da/ and /ga/ tokens is only in the third formant: the first two formants have a similar profile. This means that any changes that happen to the third formant in these tokens will lead to a higher probability of confusion between the two tokens. When CI processing is applied (ACE_{old}, ACE_{neural}, or ECC_{neural}), distortions are introduced starting from the third formant up, while most of the information in the first two formants is preserved. This may explain the larger confusion observed between the /da/ and /ga/ categories in all three processed conditions.

Another difference lies between the neural model and the standard vocoder. Whereas only the cues for /da/ and /ga/ are confused together for ACE_{old}, for ACE_{neural} all three token categories become blurred, as can be seen by the appearance of some /ba/ responses given for tokens /da-ga/ and /ga/. Such an effect can be attributed to additional smearing from the channel interaction introduced by the neural model.

Comparing the performance across ACE_{neural} and ECC_{neural} reveals that while the token /ga/ becomes more readily identified, the confusion rate for the other two categories /ba/ and /da/ increases. This may be caused by the additional spread of activity induced by the ECC coding strategy which could affect not only the third formant, but also the second.

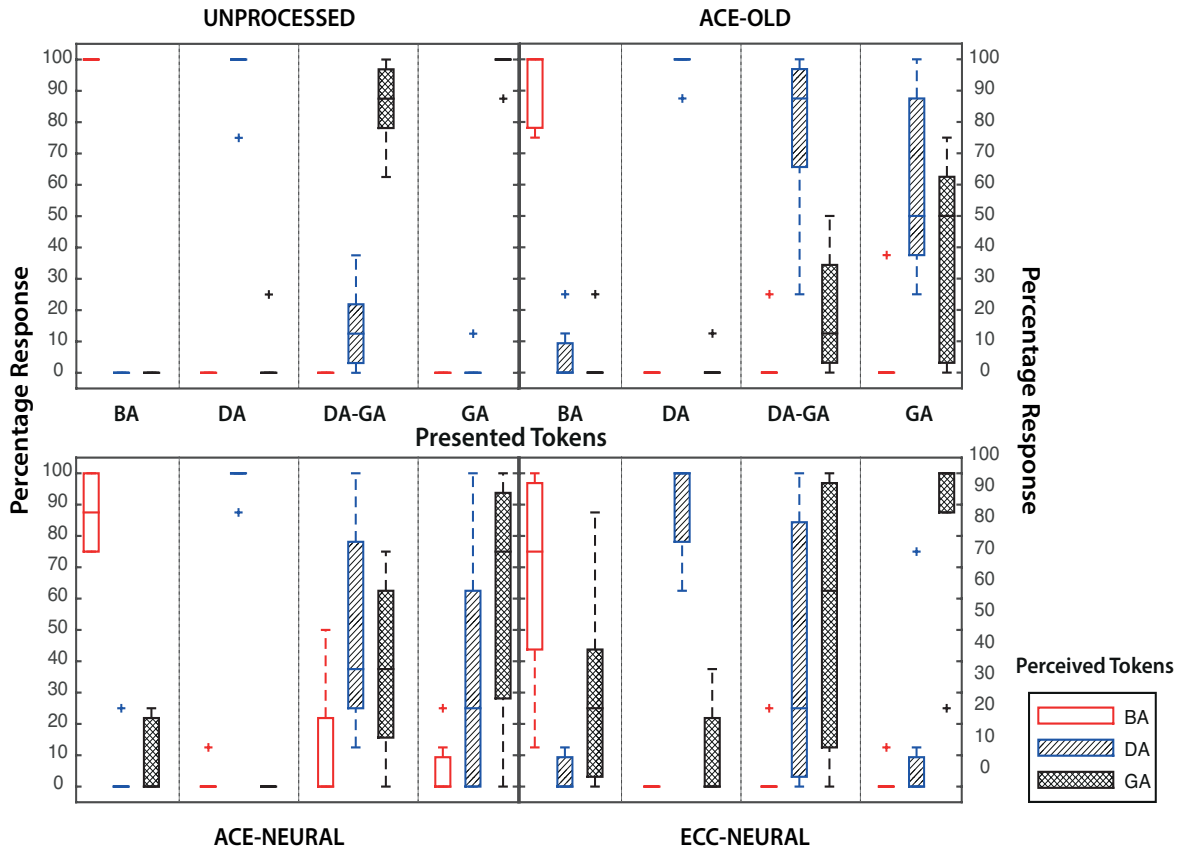


Figure 7: Plot showing how the tokens chosen for the EEG experiments are categorized in the psychophysical tests. Each box represents the percentage responses given to classify the token as /ba/ (red), /da/ (blue), or /ga/ (black).

3.3.2. VCV Tokens

The VCV tokens chosen for this experiment span sonorant (/m/, /n/, and /l/) and non-sonorant (/b/) consonant groups. /m/ and /n/ are classified as nasal stop consonants, /l/ is classified as a liquid, and /b/ as an oral stop consonant.

When ACE_{old} is used (top-right panel of Figure 9), /a-Na/ is largely perceived as /a-La/, and /a-Ma/ is sometimes confused with /a-Na/. This indicates that differences within subgroups of consonants (nasal stop) and across subgroups (nasal stop and liquids), but not across groups (sonorant and non-sonorant) are likely to be smeared.

Processing the tokens with ACE_{neural} introduces more confusion across the two subgroups nasal-stop and liquids, but not between sonorants and non-sonorants. Again this may be due to the channel interaction arising from the neural model processing.

Finally, when processing the tokens with ECC_{neural}, /a-Ma/ and /a-Na/ are largely confused with /a-La/, and /a-Ba/ becomes slightly mixed with /a-Ma/. In this case, the nasal stops and liquids are confused together in addition to some distortions that are introduced in the token /a-Ba/. The spectrum for /a-Ba/ is characterized by a visible discontinuity at the lower formants due to the narrowing of the vocal tract constriction to obstruct airflow. When ECC_{neural} is applied, some distortions are introduced in this discontinuity, thus making a smoother transition between the first and second /a/, which in turn increases the chance that /a-Ba/ is confused with other tokens that do not have a discontinuity in their spectra (like /a-Ma/).

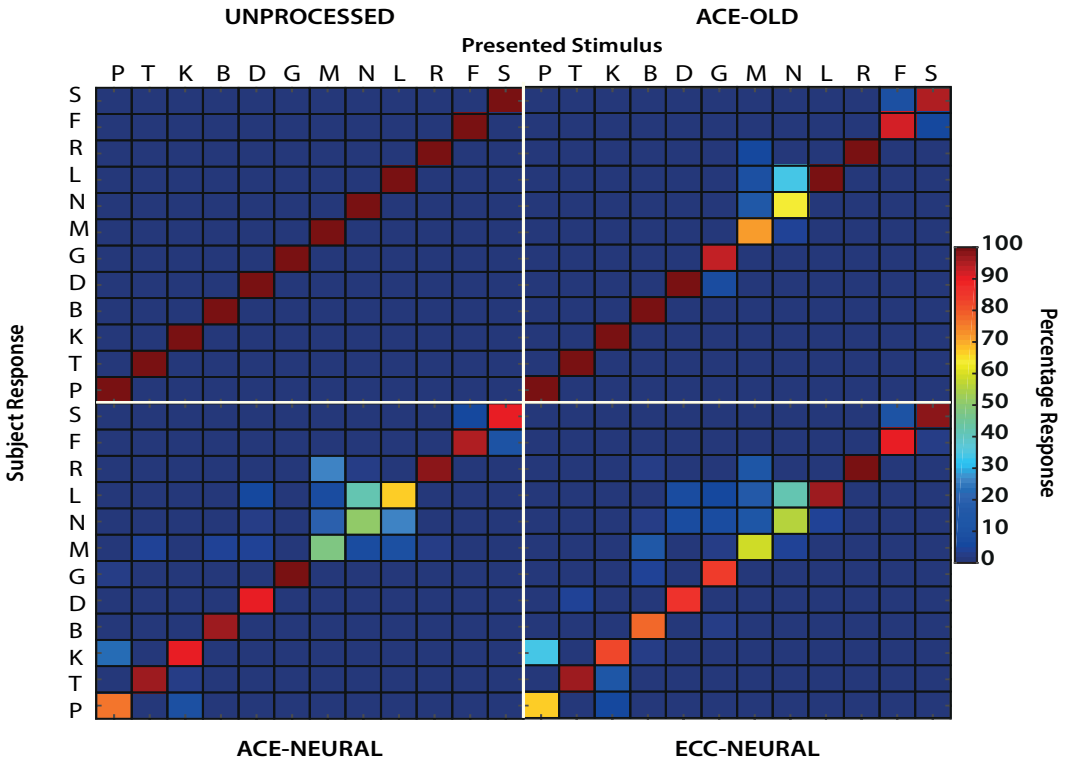


Figure 8: Confusion matrices for the VCV set.

It should be pointed out that the neural-based vocoder is not expected to yield better performance compared to the standard channel vocoder. The results from ACE_{old} are suspected to be over-optimistic since the model does not account for parasitic spread of excitation, for example, which would lead to more degraded performance. Thus, the higher confusion arising from using ACE_{neural} as opposed to ACE_{old} is in fact anticipated. However, to assess how close the neural-based vocoder comes to real CI performance would require some additional comparisons between these results from ACE_{old} and ACE_{neural} and data from actual CI patients, which is beyond the present scope of this study.

4. Experiment II: EEG Recordings

To investigate neural correlates of how similar phonemes can be distinguished from each other, MMN responses can be computed from subjects' ERP waveforms (Ortmann et al., 2013). An MMN waveform usually contains a perceptible negativity which peaks between 100 to 200 ms after the onset of a change in the stimulus (Näätänen, 2000). This can be obtained using an oddball paradigm, such as that defined in Lonka et al. (2013); Zhang et al. (2013); Kraus et al. (1993); Kühnis et al. (2013); Ortmann et al. (2013). In an oddball paradigm, two stimulus types are defined: a standard and a deviant. The standard stimulus is presented to the subjects at regular intervals and is occasionally interrupted by the deviant stimulus at pseudorandom timestamps. For example, if the standard stimulus is the speech token /da/ and the deviant is /ba/, then a typical stimulus sequence using this paradigm would be *da-da-da-...* ***ba****-da-da...****ba****-da....*, where the probability of occurrence of the standard /da/ token is usually 75-85 %, while that of the deviant /ba/ token is around 15-25% (Lonka et al., 2013; Kraus et al., 1993). The MMN waveform is then computed by subtracting the ERP response to the standard stimulus from the response to the deviant one (Näätänen, 2000).

The ERP resulting from the deviant stimulus is thought to shed some light on the perception of a difference between the standard and the deviant stimuli at the cortical level. Hence this response can be

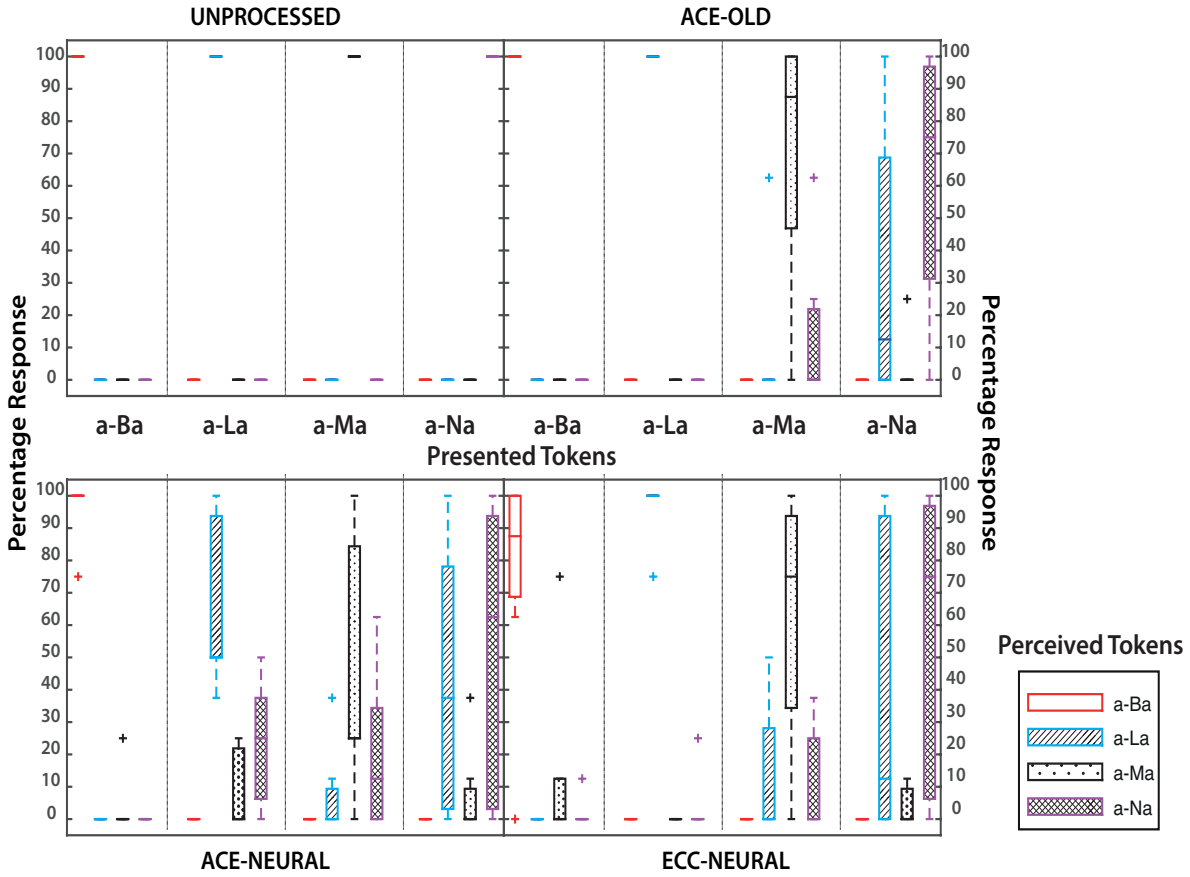


Figure 9: Plot showing how the VCV tokens chosen for the EEG experiment are categorized in the psychophysical tests. For each token, each box represents the percentage responses given to classify the token as /a-Ba/ (red), /a-La/ (blue), /a-Ma/ (black), and /a-Na/ (purple).

used as a measure for associating cortical neural responses with subjects' ability to differentiate similar phonemes, if one of the tokens is used as the standard stimulus and the other is used as the deviant (Kraus et al., 1993; Ortmann et al., 2013; Kühnis et al., 2013; Näätänen, 2000).

Because of the large number of experimental conditions tested, the *Optimum-1* oddball paradigm (Näätänen et al., 2004) was used with some modifications. The stimulus stream began with 15 presentations of the standard followed by alternating deviant-standard pairs, which were put in groups containing each of the 3 deviants tested. In each repetition of the group, the deviants were pseudorandomly shuffled.

There were 40 presentations per deviant and 130 presentations of the standard token for each of the two stimulus groups under each of the four experimental conditions. All stimuli were delivered using the software Presentation (NeuroBehaviouralSystems, http://www.neurobs.com). EEG responses were recorded for each subject and stored for further offline processing.

4.1. Procedure

Subjects were seated in an electromagnetically-shielded room, in which they were stimulated in the right ear via an insert earphone while the left ear was plugged. They were asked to watch a silent subtitled movie as a distractor task (Lonka et al., 2013; Zhang et al., 2013; Kraus et al., 1993; Kühnis et al., 2013; Ortmann et al., 2013).

Since the goal of this experiment was to examine responses to easy and difficult token contrasts, three token contrasts were taken from the two stimulus groups as follows.

4.1.1. CV Tokens

The anchor token /da/ was chosen as the standard stimulus with three different contrast categories: the anchor token /ba/ represents an easy contrast because it is easily distinguishable from /da/ under the different processing conditions. The anchor token /ga/ was chosen to represent a more difficult contrast, while the middle token between the /da/ and /ga/ transitions was chosen to represent the most difficult case of the three, in which the deviant cannot be distinguished from the standard even in the unprocessed case. This token is referred to as *intermediate /da-ga/* in the rest of this paper. The easy contrast is expected to elicit the largest MMN waveform compared to the other two cases, while the response to the intermediate /da-ga/ token is expected to be the smallest.

4.1.2. VCV Tokens

For this set, /a-La/ was chosen as the standard, and the contrast /a-La/ versus /a-Ba/ was taken as the easy contrast because those two tokens are seldom confused with one another under the different processing conditions. The contrasts /a-La/ vs. /a-Ma/ and /a-La/ vs. /a-Na/ were chosen as difficult contrasts because they are confused together across three of the four experimental conditions (see Figure 9).

4.2. EEG Data Analysis

The actiCAP 64Ch Standard-2 electrode cap configuration, which is a 64-electrode setup numbered according to the 10-20 system, was used. Electrode impedances were kept below 30 KΩ using the *actiCAP* software.

All EEG recordings were captured via Brain Vision Recorder. Waveforms were bandpass-filtered between 0.1-250Hz, and sampled at a rate of 500 Hz. The amplifier resolution was set to 0.1 μV, with positive polarity set to point upwards.

The recorded EEG data were processed offline using Brain Vision Analyzer 2.1. EEG waveforms were filtered using a Butterworth zero phase filter, with a low cutoff of 1 Hz at 12 dB/octave, a high cutoff of 25 Hz at 24 dB/octave, and a notch filter at 50 Hz to get rid of interference from electrical equipment in the room.

The Infomax Independent Component Analysis (ICA) algorithm was used to extract eye-blinks and horizontal eye movement artefacts. The number of ICA components was set to the maximum number of channels (64).

Data was re-referenced to the average of the two mastoid channels TP9 and TP10 to ensure that the MMN waveform is unbiased towards either of the hemispheres (Luck, 2005).

Epochs were taken starting 50 ms pre-stimulus onset till 500 ms post-stimulus onset, were corrected locally for DC drifts, and were baseline-corrected starting at 50 ms pre-stimulus onset. For CV tokens, a single trigger was placed at stimulus onset, while for the VCV tokens two triggers were used. The first trigger was placed at stimulus onset, while the second was placed approximately where the consonant starts.

MMN waveforms were calculated for N1 peaks, defined in the region between 100 and 200 ms post-stimulus onset.

4.3. Results

ERP responses to the deviants were rather weak compared to the data provided in Ortmann et al. (2013); Sandmann et al. (2009, 2010), and thus the results reported here are provided to help visualise the differences between those waveforms in the region between 100 ms and 200 ms.

4.3.1. CV Tokens

Figure 10 shows the overlaid MMN responses for each token contrast under each of the four experimental conditions.

From the psychophysical data for the unprocessed condition (top-left panel in Figure 7), it can be expected that MMN waveforms should contain a discernible peak for the deviants /ba/ and /ga/, but not for intermediate /da-ga/, since it is highly confused with /da/. However, the MMN waveforms in Figure 10(a) elicited by the deviants /ga/ and intermediate /da-ga/ fail to show any noticeable negativities in the time between 100 and 200 ms.

The psychophysical data in Figure 7 for ACE_{old} indicate that while the token /ba/ may be clearly distinguished from /da/, the tokens /ga/ and intermediate /da-ga/ may not be as easily identified. This is due to the higher confusion rate with /da/ in contrast to the unprocessed condition. From that data, it can be speculated that the MMN peak for /ba/ should be much larger than that elicited by /ga/ or intermediate /da-ga/. However, the difference waves extracted for ACE_{old} revealed positive mismatch peaks of comparable amplitudes for all three deviants (Figure 10(b)).

For the ACE_{neural} condition, the psychophysical data indicate that the cues for distinguishing /ga/ from /da/ become more salient compared to ACE_{old}. This means that the MMN peak resulting from /ga/ and intermediate /da-ga/ should be larger than those elicited by the same contrasts in the ACE_{old} scenario. Examining the MMN waveforms in Figure 10(c) fails to reveal discernible negativities for the deviants /ba/ and /ga/.

Finally, for the ECC_{neural} condition, the behavioural data show that the subjects were able to distinguish /ga/ from /da/ at a much higher rate compared to the previous two processing conditions that relied on ACE. On the other hand, the cues for /ba/ appear to have been smeared as can be inferred from the large variability obtained for this token (Figure 7, bottom-right panel). This means that the MMN peak for /ga/ should be larger than that for /ba/, which should in turn be larger than that for intermediate /da-ga/. Figure 10(d) reveals this to be the case, because for this experimental condition the token /ga/ was the least confused with /da/ as can be seen in the behavioural responses in Figure 7.

4.3.2. VCV Tokens

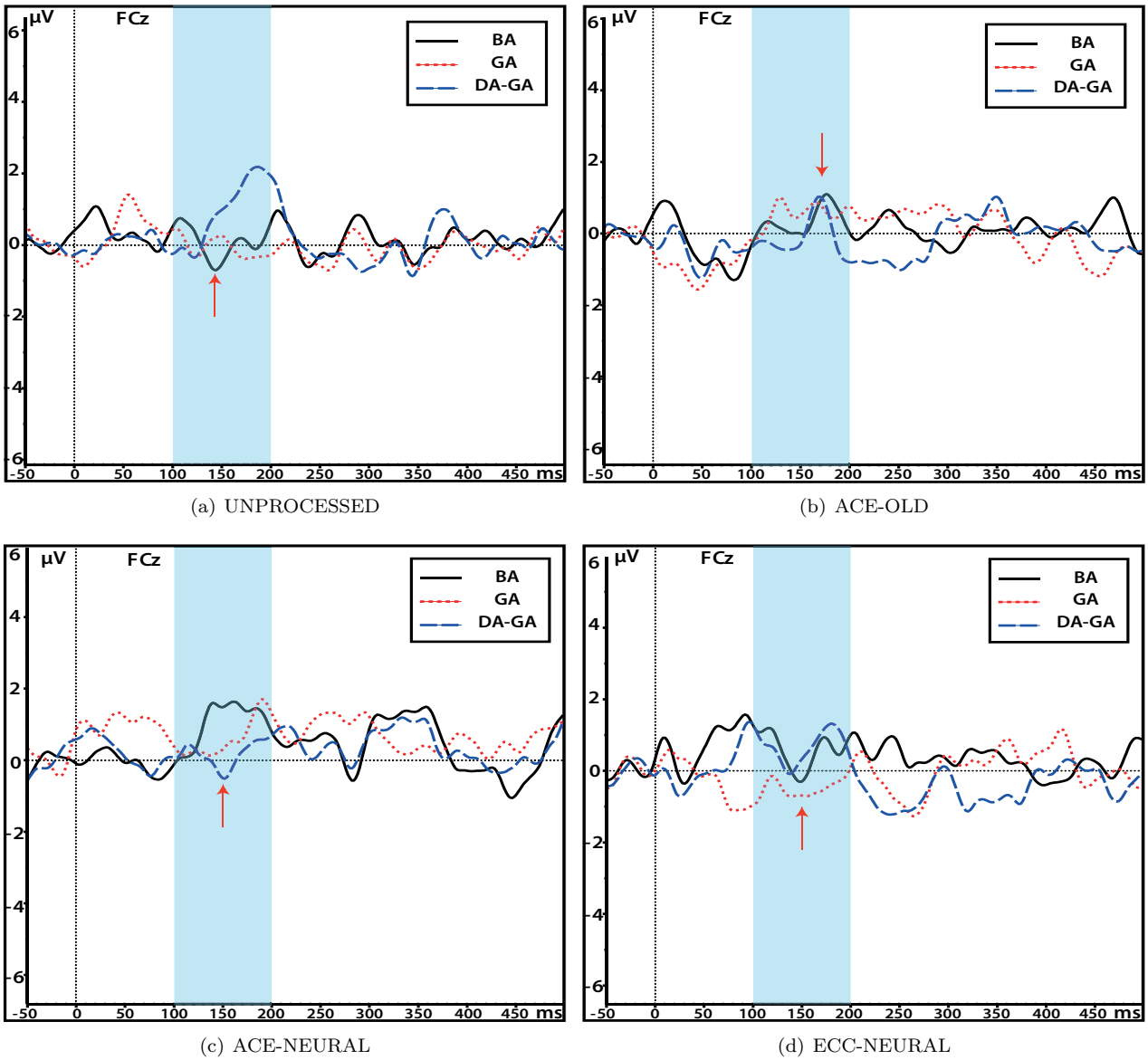
Results for the VCV tokens are shown in Figure 11. Examining the psychophysical data in Figure 9 reveals that for the unprocessed case, all tokens are very clearly identified. This means that the MMN responses for all three tokens (/a-Ba/, /a-Ma/, and /a-Na/) should be comparable. Figure 11(a) does show similar waveforms, except that the MMN response for /a-Ma/ is larger than that of /a-Ba/, which is in turn larger than the response to /a-Na/.

The ACE_{old}, ACE_{Neural}, and ECC_{neural} panels in Figure 9 show that the tokens /a-Ma/ and /a-Na/ are largely perceived as /a-La/ in the CI simulations especially in the case where the neural-based vocoder (ACE or ECC) is used. This is expected to cause ERPs in response to both /a-Ma/ and /a-Na/ to be very similar to that of /a-La/, with $ERP_{/a-Ma/}$ being slightly larger than $ERP_{/a-Na/}$. Thus the response to the easy contrast /a-Ba/ is expected be the largest of the three. Figures 11(b) and 11(c) show this expected behaviour for ACE_{old} and ACE_{neural}, respectively. However, for the ECC condition (Figure 11(d)), this trend is lost since all deviants yield almost the same MMN amplitude.

4.4. Discussion

Although the aforementioned results seem to indicate that a trend that supports the posited hypothesis may exist, solid conclusions cannot be drawn regarding whether this trend in MMN responses is truly robust enough to be manifest in a larger number of participants.

These results raise the question of whether the Optimum-1 paradigm was in fact suitable for testing the hypothesis in this study. The paradigm was introduced in Näätänen et al. (2004) with tone stimuli and not with speech. In this sense, it may be that using this paradigm with speech stimuli does not reveal robust MMN peaks. This idea was tested by averaging the ERPs from all deviants across all participants for the unprocessed condition and comparing that waveform to the grand-average ERP obtained from the standard token. The ERP obtained for all the deviants averaged together was smaller than that obtained for the



(a) UNPROCESSED

(b) ACE-OLD

(c) ACE-NEURAL

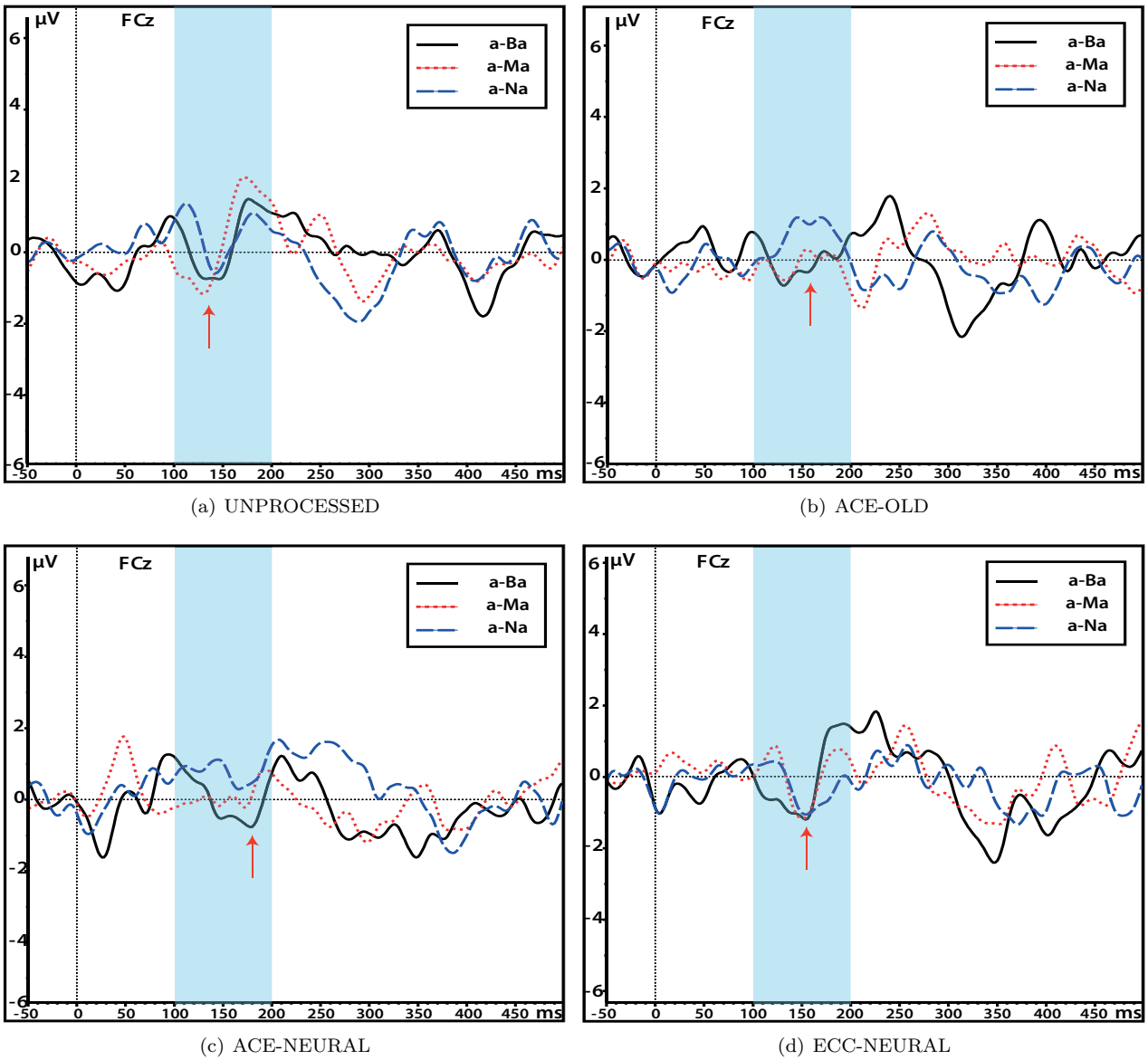
(d) ECC-NEURAL

Figure 10: Grand Average MMN waveforms for the CV tokens. The zero point represents stimulus onset. The black curves denote the MMN response for the deviant /ba/, the red curves are the MMN responses for /ga/, and the blue curves are for the intermediate /da-ga/ token.

standard, which indicates that the paradigm might not have worked with the current speech stimuli. In that case, the traditional oddball paradigm might have provided better results.

Additionally, the nature of the stimuli might have had an affect on the MMN waveforms seen. Using vocoded speech tokens may have induced smaller ERPs for the deviants because the stimuli now sound unnatural to the NH participants. Thus small differences between the tokens could no longer be consciously detected, especially since the duration of the consonant part of the CV tokens was quite short (40 ms long).

For the VCVs, trigger placement was quite challenging because the beginning of the response to the consonant is unknown. Since the consonant occurs roughly between 150 ms and 180 ms post-stimulus onset, the actual response to that consonant is expected to occur between 280 ms and 380 ms instead of in the



(a) UNPROCESSED

(b) ACE-OLD

(c) ACE-NEURAL

(d) ECC-NEURAL

Figure 11: Grand Average MMN waveforms for the VCV tokens. The zero point represents stimulus onset. The black curves denote the MMN response for the deviant /a-Ba/, the red curves are the MMN responses for /a-Ma/, and the blue curves are for the /a-Na/ token.

range 100 ms - 200 ms. For the unprocessed condition, the MMN peak between 280 ms and 380 ms is largest for /a-Na/ compared to that /a-Ma/, which is in turn larger than that of /a-Ba/. In fact the response to /a-Ba/ is quite small, which does not correlate with the psychophysics. For all CI processing conditions tested, the peak for /a-Ba/ becomes quite prominent in that time range, with the responses to /a-Ma/ and /a-Na/ being comparatively smaller. However, because these responses may be influenced by the response to first /a/ in the token, no solid conclusions can be drawn from these trends. Additionally, since the VCV tokens used in this study are naturally spoken, it is difficult to compare them because they are not strictly consistent in terms of duration and emphasis.

Moreover, only 40 presentations per deviant were delivered, as was done in Henkin et al. (2008); Sand-

mann et al. (2010), to save significant testing time per participant. However, these are probably not enough and thus one option would be either to increase the number of presentations per deviant, or to increase the number of participants to guarantee that enough segments are present for averaging.

5. Conclusion

The neural model presented here allows not only simulating the output of the custom-designed ECC strategy, but also that of other newly developed coding strategies. This is beneficial in the sense that the output pulse sequences generated by experimental coding strategies can be resynthesized and subjectively evaluated, which can assist designers in the development phase. Moreover, because the neural-based vocoder incorporates characteristics of the neural interface, it may serve as a more accurate simulation compared to standard channel vocoders.

Incorporating more complex functionality, such as inhibition and adaptation, may also influence the output of the synthesizer. Additionally, the spread of activity across electrodes should be explored as a function of the electrode location (basal vs. apical) as opposed to using the same current spread profile for all electrodes. The size and location of neuron populations should not be restricted to electrode positions, which means that maximum stimulation may fall anywhere along the neural population. This is in contrast to having maximum stimulation at the centre of the population as is done in the proposed model.

Moreover, the model parameters need to be tested with unilaterally implanted CI subjects who have sufficient residual hearing in the non-implanted ear. The procedure would be to stimulate the non-implanted ear with the output of the neural-based vocoder, while stimulating the implanted ear with the unprocessed version of the same stimulus and asking subjects to subjectively rank the two stimuli. Additionally, psychophysical tests can be carried out using this stimulation method to assess how close the neural-based vocoder simulations are to capturing the cues transmitted by the implant.

While the performance of the NH participants differs substantially across vocoding conditions, this dataset needs to be compared to one acquired from CI subjects listening to ACE and ECC before any meaningful comparisons between the two coding strategies, or between the standard and neural-based vocoder implementations can be made.

In spite of the fact that the data acquired from the EEG experiments failed to give much information regarding expected trends in MMN responses to varying difficulties of speech tokens, the outcome of this pilot study serves as a first step in formulating a well-defined procedure for conducting a larger-scale study.

6. Acknowledgements

The authors would like to thank Sonia Tabibi and Dietmar Wohlbauer for their assistance with the EEG experiments. The authors would also like to thank Nathalie Giroud and Anita Wagner for their input on the EEG data analyses.

The work presented herein was part of the first author's master thesis, and was funded by The Eidgenössische Stipendienkommission für Ausländische Studierende (ESKAS), Switzerland in the form of a 1.5-year master program scholarship.

References

Adamson, C. L., Reid, M. A., and Davis, R. L. (2002). Opposite actions of brain-derived neurotrophic factor and neurotrophin-3 on firing features and ion channel composition of murine spiral ganglion neurons. *The Journal of neuroscience*, 22(4):1385–1396.

Agrawal, D., Thorne, J., Viola, F., Timm, L., Debener, S., Büchner, A., Dengler, R., and Wittfoth, M. (2013). Electrophysiological responses to emotional prosody perception in cochlear implant users. *NeuroImage: clinical*, 2:229–238.

Babacan, O., Lai, W. K., Killian, M., and Dillier, N. (2010). Implementation of a neurophysiologically-based coding strategy for the cochlear implant. In *13. Jahrestagung der Deutschen Gesellschaft für Audiologie. DGA e.V., Frankfurt*, pages 1–4. ISBN 978-3-9813141-0-6.

Başkent, D. and Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners. *The Journal of the Acoustical Society of America*, 116(5):3130–3140.

Bruce, I. C., Irlicht, L. S., White, M. W., O'Leary, S. J., Dynes, S., Javel, E., and Clark, G. M. (1999a). A stochastic model of the electrically stimulated auditory nerve: pulse-train response. *Biomedical Engineering, IEEE Transactions on*, 46(6):630–637.

Bruce, I. C., White, M. W., Irlicht, L. S., O'Leary, S. J., and Clark, G. M. (1999b). The effects of stochastic neural activity in a model predicting intensity perception with cochlear implants: low-rate stimulation. *Biomedical Engineering, IEEE Transactions on*, 46(12):1393–1404.

Bruce, I. C., White, M. W., Irlicht, L. S., O'Leary, S. J., Dynes, S., Javel, E., and Clark, G. M. (1999c). A stochastic model of the electrically stimulated auditory nerve: single-pulse response. *Biomedical Engineering, IEEE Transactions on*, 46(6):617–629.

Büchner, A., Nogueira, W., Edler, B., Battmer, R.-D., and Lenarz, T. (2008). Results from a psychoacoustic model-based strategy for the nucleus-24 and freedom cochlear implants. *Otology & Neurotology*, 29(2):189–192.

Chatterjee, M. and Zwislocki, J. J. (1998). Cochlear mechanisms of frequency and intensity coding. ii. Dynamic range and the code for loudness. *Hearing research*, 124(1):170–181.

Chen, F. (2012). Predicting the intelligibility of cochlear-implant vocoded speech from objective quality measure. *J. Med. Biol. Eng*, 32:189–194.

Chen, F. and Loizou, P. C. (2011). Predicting the intelligibility of vocoded speech. *Ear and hearing*, 32(3):331.

Chen, F. and Zhang, Y.-T. (2007). An integrate-and-fire-based auditory nerve model and its response to high-rate pulse train. *Neurocomputing*, 70(4):1051–1055.

Cohen, L. T. (2009a). Practical model description of peripheral neural excitation in cochlear implant recipients: 1. growth of loudness and ECAP amplitude with current. *Hearing research*, 247(2):87–99.

Cohen, L. T. (2009b). Practical model description of peripheral neural excitation in cochlear implant recipients: 2. spread of the effective stimulation field (esf), from ECAP and FEA. *Hearing research*, 247(2):100–111.

Cohen, L. T. (2009c). Practical model description of peripheral neural excitation in cochlear implant recipients: 3. ECAP during bursts and loudness as function of burst duration. *Hearing research*, 247(2):112–121.

Cohen, L. T. (2009d). Practical model description of peripheral neural excitation in cochlear implant recipients: 4. model development at low pulse rates: General model and application to individuals. *Hearing research*, 248(1):15–30.

Cohen, L. T. (2009e). Practical model description of peripheral neural excitation in cochlear implant recipients: 5. refractory recovery and facilitation. *Hearing research*, 248(1):1–14.

Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. S. (2003). Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ECAP method and psychophysical forward masking. *Hearing research*, 179(1):72–87.

Digeser, F. M., Wohlberedt, T., and Hoppe, U. (2009). Contribution of spectrotemporal features on auditory event-related potentials elicited by consonant-vowel syllables. *Ear and hearing*, 30(6):704–712.

Dorman, M. F., Loizou, P. C., and Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *The Journal of the Acoustical Society of America*, 102(4):2403–2411.

Florentine, M. and Zwicker, E. (1979). A model of loudness summation applied to noise-induced hearing loss. *Hearing research*, 1(2):121–132.

Fredelake, S. (2012). *Model-based prediction of the benefit with rehabilitative hearing devices*. PhD thesis, Universität Oldenburg.

Fredelake, S. and Hohmann, V. (2012). Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation. *Hearing research*, 287(1):76–90.

Fu, Q.-J. (2005). Loudness growth in cochlear implants: effect of stimulation rate and electrode configuration. *Hearing research*, 202(1):55–62.

Fu, Q.-J. and Shannon, R. V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *The Journal of the Acoustical Society of America*, 105(3):1889.

Gerstner, W. and Kistler, W. M. (2002). *Spiking Neural Models: Single Neurons, Populations, Plasticity*, chapter 6. Population Equations. Cambridge University Press.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species29 years later. *The Journal of the Acoustical Society of America*, 87(6):2592–2605.

Hamacher, V. (2004). *Signalverarbeitungsmodelle des elektrisch stimulierten Gehrs; 1. Aufl.* PhD thesis, Aachen. Zugl.: Aachen, Techn. Hochsch., Diss., 2003.

Hant, J. J. and Alwan, A. (2000). Predicting the perceptual confusion of synthetic stop consonants in noise. In *ICSLP*, volume 3, pages 941–944.

Hartmann, R., Topp, G., and Klinke, R. (1984). Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea. *Hearing research*, 13(1):47–62.

Henkin, Y., Tetin-Schneider, S., Hildesheimer, M., and Kishon-Rabin, L. (2008). Cortical neural activity underlying speech perception in postlingual adult cochlear implant recipients. *Audiology and Neurotology*, 14(1):39–53.

Holmberg, M., Gelbart, D., and Hemmert, W. (2007). Speech encoding in a model of peripheral auditory processing: Quantitative assessment by means of automatic speech recognition. *Speech Communication*, 49(12):917–932.

Iverson, P. (2001). Individual differences in phonetic perception by adult cochlear implant users: Effects of sensitivity to/d/-/t/on word recognition. *Speech, Hearing, and Language: work in progress*, 13:1–21.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America*, 67(3):971–995.

Kraus, N., Micco, A. G., Koch, D. B., McGee, T., Carrell, T., Sharma, A., Wiet, R. J., and Weingarten, C. Z. (1993). The mismatch negativity cortical evoked potential elicited by speech in cochlear-implant users. *Hearing research*, 65(1):118–124.

Kühnis, J., Elmer, S., Meyer, M., and Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: an eeg study. *Neuropsychologia*, 51(8):1608–1618.

Lai, W. K., Bögli, H., and Dillier, N. (2003). A software tool for analyzing multichannel cochlear implant signals. *Ear and hearing*, 24(5):380–391.

Lai, W. K. and Dillier, N. (2002). Macarena: a flexible computer-based speech testing environment. In *Proceedings of the 7th International Cochlear Implant Conference*.

Lai, W. K. and Dillier, N. (2012). Encoding signal intensity with stimulation rate: Forward masking measures. In *15. Jahrestagung der Deutschen Gesellschaft für Audiologie. DGA e.V., Erlangen*, pages 1–4. ISBN 978-3-9813141-2-0.

Lai, W. K. and Dillier, N. (2013). Rfcap: A software analysis tool for multichannel cochlear implant signals. *Cochlear implants international*, 14(2):107–116.

Laneau, J., Moonen, M., and Wouters, J. (2006). Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. *The Journal of the Acoustical Society of America*, 119(1):491–506.

Lonka, E., Relander-Syrjänen, K., Johansson, R., Näätänen, R., Alho, K., and Kujala, T. (2013). The mismatch negativity (mmn) brain response to sound frequency changes in adult cochlear implant recipients: a follow-up study. *Acta oto-laryngologica*, 133(8):853–857.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. The MIT Press.

Massida, Z., Belin, P., James, C., Rouger, J., Fraysse, B., Barone, P., and Deguine, O. (2011). Voice discrimination in cochlear-implanted deaf subjects. *Hearing research*, 275(1):120–129.

McKay, C. M., Henshall, K. R., Farrell, R. J., and McDermott, H. J. (2003). A practical method of predicting the loudness of complex electrical stimuli. *The Journal of the Acoustical Society of America*, 113(4):2054–2063.

McKay, C. M. and McDermott, H. J. (1998). Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals. *The Journal of the Acoustical Society of America*, 104(2):1061–1074.

Miller, C. A., Abbas, P. J., and Robinson, B. K. (2001). Response properties of the refractory auditory nerve fiber. *Journal of the Association for Research in Otolaryngology*, 2(3):216–232.

Näätänen, R. (2000). Mismatch negativity (mmn): perspectives for application. *International Journal of Psychophysiology*, 37(1):3–10.

Näätänen, R., Pakarinen, S., Rinne, T., and Takegata, R. (2004). The mismatch negativity (mmn): towards the optimal paradigm. *Clinical Neurophysiology*, 115(1):140–144.

Negm, M. H. and Bruce, I. C. (2014). The Effects of HCN and KLT Ion Channels on Adaptation and Refractoriness in a Stochastic Auditory Nerve Model. *Biomedical Engineering, IEEE Transactions on*, 61(11):2749–2759.

Nogueira, W., Büchner, A., Lenarz, T., and Edler, B. (2005). A psychoacoustic nofm-type speech coding strategy for cochlear implants. *EURASIP Journal on Applied Signal Processing*, 2005:3044–3059.

Omran, S. A., Lai, W. K., and Dillier, N. (2010). Pitch ranking, melody contour and instrument recognition tests using two semitone frequency maps for nucleus cochlear implants. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010:13.

Ortmann, M., Knief, A., Deuster, D., Brinkheetker, S., Zwitserlood, P., am Zehnhoff-Dinnesen, A., and Dobel, C. (2013). Neural correlates of speech processing in prelingually deafened children and adolescents with cochlear implants. *PloS one*, 8(7):e67696.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., Chang, E. F., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS-Biology*, 10(1):175.

Raz, I. and Noffsinger, D. (1985). Identification of synthetic, voiced stop-consonants by hearing-impaired listeners. *International Journal of Audiology*, 24(6):437–448.

Sandmann, P., Eichele, T., Buechler, M., Debener, S., Jäncke, L., Dillier, N., Hugdahl, K., and Meyer, M. (2009). Evaluation of evoked potentials to dyadic tones after cochlear implantation. *Brain*, 132(7):1967–1979.

Sandmann, P., Kegel, A., Eichele, T., Dillier, N., Lai, W. K., Bendixen, A., Debener, S., Jancke, L., and Meyer, M. (2010). Neurophysiological evidence of impaired musical sound perception in cochlear-implant users. *Clinical Neurophysiology*, 121(12):2070–2082.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234):303–304.

Sotscheck, J. (1982). A rythme test for intelligibility measurements in german as an improved procedure for determining speech transmission quality. *Speech Communication*, 1(1):83.

Stakhovskaya, O., Sridhar, D., Bonham, B. H., and Leake, P. A. (2007). Frequency map for the human cochlear spiral ganglion: implications for cochlear implants. *Journal for the Association for Research in Otolaryngology*, 8(2):220–233.

Stephens, J. D. and Holt, L. L. (2011). A standard set of American-English voiced stop-consonant stimuli from morphed natural speech. *Speech communication*, 53(6):877–888.

Strydom, T. and Hanekom, J. J. (2011). The performance of different synthesis signals in acoustic models of cochlear implants. *The Journal of the Acoustical Society of America (JASA)*, 129(2):920–933.

Waters, J. and Helmchen, F. (2006). Background synaptic activity is sparse in neocortex. *The Journal of neuroscience*, 26(32):8267–8277.

Yates, G. K., Winter, I. M., and Robertson, D. (1990). Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. *Hearing research*, 45(3):203–219.

Zeng, F.-G., Rebscher, S., Harrison, W., Sun, X., and Feng, H. (2008). Cochlear implants: system design, integration, and evaluation. *Biomedical Engineering, IEEE Reviews in*, 1:115–142.

Zhang, F., Benson, C., and Fu, Q.-J. (2013). Cortical encoding of pitch contour changes in cochlear implant users: A mismatch negativity study. *Audiology and Neurotology*, 18(5):275–288.

Zwislocki, J. (1973). On intensity characteristics of sensory receptors: A generalized function. *Kybernetik*, 12(3):169–183.